“ Grab ’em by the Fallacy ”
Stance Detection to Identify Fake News

CS585, UMass Amherst, Fall 2017

Ajinkya Zadbuke **Arhum Savera**
azadbuke@cs.umass.edu asavera@umass.edu

1 Outline

With the constant deluge of information available through multiple sources, it is more important than ever to be able to distinguish between factual and fabricated reports. Fake news, defined by the New York Times as ’a made up story with an intention to deceive’ has been widely cited as a contributor to the outcome of the 2016 US elections. The Fake News Challenge was organized as a response to this controversy [1]. It is a public competition with the goal of using AI to automatically identify fake news. In this project, we aim to implement a solution for detecting stance in news articles, comparing multiple approaches and models. Our primary interest is in applying deep learning architectures such as Feed-forward Neural Networks, RNNs and LSTMs for this problem.

2 Literature Review

Paper [2] experiments with Conditional LSTM encoding in context of a tweet-to-target model, with a slightly different labelling scheme. It involves encoding the target as a fixed-length vector, and then the tweet with state intialized as the target’s repesentation. The final output of the tweet LSTM is used to predict the stance label for the pair.
Paper [3] tackles the same fake news dataset that we are using, implementing a Concatenated Multi-Layer Perceptron model, with Bag of Words representations and GLoVE Embeddings. They have also tried RNN and LSTM models, with the MLP performing better than the others.
Paper [4] also tackles the problem of fake news detection but uses a combination of linguistic cue approaches (with machine learning) and network analysis approaches. They experimented with BoW, Probabilistic Context-Free Grammars and SVMs while taking Social Network behavior and linked data properties into account.

3 Datasets

The dataset we will use is the one provided for the original FNC task [5]. This is a collection of articles, each consisting of a (headline, body, stance). The stance, which is our label, in this case, is one of {unrelated, disagree, agree, discuss}. The dataset has been created and annotated by accredited journalists, so we are not concerned with verifying quality of the data. We have 1648 headlines and 1669 article bodies, which have paired to create 49972 body-headline pairs. The data is inherently imbalanced as a result of the pairwise combinations, such that around 73%

of the examples are labeled as 'unrelated' with the rest distributed between 'discuss', 'agree' and 'disagree'. We may decide to use additional data from other sources later.

4 Evaluation

Keeping with the spirit of the challenge, we will follow their evaluation guidelines. This involves handling the class imbalance by assigning a 25-75 weighting for the 'unrelated' and other classes, respectively. The baseline accuracy provided by the challenge can serve as a good benchmark to judge our models against (currently at 79.53%).

5 Scope and Approach

The scope of the project is constrained by the following. Claims made by recognized newsworthy sources are not considered fake news, as well as articles with humorous or sarcastic intent, as opposed to fabricated claims about newsworthy people, or by disreputable sources. The scope is also constrained by the data we use. Given the data we have, we focus on stance detection for the above, rather than explicitly modelling the fake news problem. Truth labelling is a challenging task even for humans, and other tasks have been determined to be feasible but decidedly non-trivial [5]. Our approach will involve implementing increasingly complex models, starting from a simple Logistic Regression classifier, and moving on to deep neural architectures such as Recurrent Neural Networks and Long-Short-Term Memory Networks. We aim to compare and contrast the performance of these models on a development set to tune our model hyperparameters (using k-fold cross-validation), and evaluate final accuracy on the test set. The test set itself consists of around 1000 samples, with the distribution of the classes unknown.

6 Pre-existing Software Systems that can be used

(a) Implementation of code : Python
(b) Natural Language Processing over dataset : SpaCy, NLTK
(c) Machine Learning Libraries : Scikit-Learn, PyTorch/Tensorflow
(d) General purpose computation and visualization : Numpy, Pandas, Jupyter, Matplotlib.

7 References

[1] Fake News Challenge - https://www.fakenewschallenge.org/
[2] Augenstein, I., Rocktaschel, T., Vlachos, A., Bontcheva, K., *Stance Detection with Conditional Bidirectional Encoding*, September 2016
[3] *Fake News, Real Consequences: Recruiting Neural Networks for the Fight Against Fake News*, January 2017
[4] Conroy, N., Rubin, V., Chen, Y. *Automatic Deception Detection: Methods for Finding Fake News*, November 2015
[5] Ferreira, W., and Vlachos, A., *Emergent : A Novel Dataset for Stance Classification*, June 2016